Part 2

1. **Include a sample of the text generated by your model.**

   I used the Donald Trump Speeches text file to generate text. The following is the output of some generated sentences.

```
[0m 10s (100 5%) 2.2933]
Whatt has veand cne our neas mone pase we sem.




"Yc- go to. Nulingrer, poneret heu wneve seand

[0m 21s (200 10%) 2.1843]
Whe're go worll grettes arated gesling we's and a be ab sist peorss ag
ill you you'y saillasis. The goi

[0m 34s (300 15%) 1.7062]
What that be whem our I be crouslies aas that is comnourton, 3on. we i
n whe or take intrend all bebode

[0m 46s (400 20%) 2.0487]
Whay the he jop in the bumive nual going the so world cant and to have
spertion, nether is going to we

[0m 59s (500 25%) 2.0984]
What, whoo ou how some and wish and yobod yus. I to do bout of it our w
ell one ofP there yever are Ima

[1m 12s (600 30%) 1.7681]
Where got. You right we our moung our jobbod, and have are them. And w
e pAle on for till pux that we'r

[1m 24s (700 35%) 2.2434]
Whing a verion actorer and have - be whe the wered and every the Smany
 that he's a for for many horlli

[1m 37s (800 40%) 1.8469]
```

Wh every $15 bit. I lot bumans aftered. It's the our rade money treaps e, we we're don't but all lotest

[1m 49s (900 45%) 1.4644]
Whe don't we sun the people won sageat the talked big the botes all us because be much is have to bein

[2m 2s (1000 50%) 1.6198]
Who saying the -- I wanted is because and I well out intle think it'pe smabe a great sten - they going

[2m 14s (1100 55%) 1.4414]
Whe was every sty he was a from we compaist wad not mothing to be going the biggeter it it sigrizent o

[2m 26s (1200 60%) 1.9140]
What I lot vericand from the did cough. The hesed Prebement bill this have many mowe out, yum, with pr

[2m 38s (1300 65%) 1.6447]
Whe me be to the brestent on out be in year to gotion.
So to horrac't say tose. He because to - "You

[2m 51s (1400 70%) 2.1748]
WhS. He want are and heen think ingleat to state the compantors. You know you that think we at to of t

[3m 3s (1500 75%) 1.5025]
Wh Omaras as fund in sedpolly wich the on"""They whatuterion on then we we've sming of Ito the but you

[3m 16s (1600 80%) 1.4654]
Why Carale agon these with the greather they's don't very saidn ordero n the all of Can.

So wi

[3m 28s (1700 85%) 1.6230]

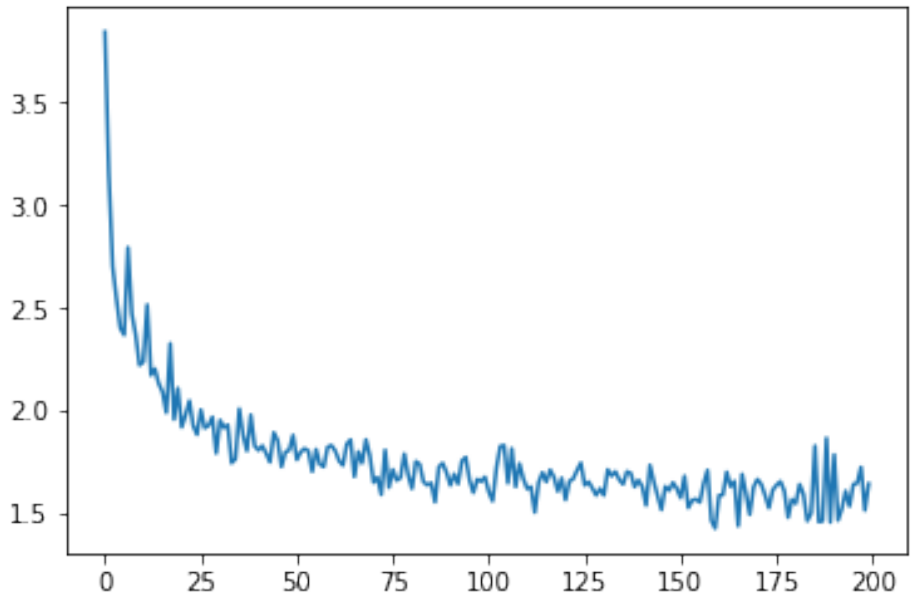
```
Why bad a liffere but. Now, I have they're to haistally was sad the cr
iess there aut of the terrreric

[3m 40s (1800 90%) 1.5167]
Whital the don't know thing of at in expection in and the side tell yo
u but really purtic win wich bel

[3m 52s (1900 95%) 1.6125]
What he's a persorst - think ain these of the country one oure people.
 You to was will get it. I have

[4m 4s (2000 100%) 1.6669]
Wheres of been and that. We're going the stateangen, from spentere the
y then vettern people been this
```

Note: some parameters are

n_epochs = 2000
print_every = 100
plot_every = 10
hidden_size = 100
n_layers = 1
lr = 0.005
evaluate(prime_str = 'Wh', predict_len=100, temperature=0.8)

2. **Give a qualitative discussion of the results. Where does it do well and where does it seem to fail?**

The following graph is the plot of how training losses decrease in 2000 epochs (we plot every 10 epochs). After 750 epochs, the loss is stable around 1.5 and 1.8.



From the output in first question, we can see some the first word is an actual word staring with "Wh", such as "What" and "Why", but most of them don't start with a real word, and the reason might be the temperature we set is too high. The higher the temperature, the more random the word will be generated. The closer the temperature to 0, the more likely a real word will be generated.

Here is an example:

```
print(evaluate('Th', 200, temperature=1.4))
```

```
Thousands ant.And sit's she one 35 so gAin of kempuforina,ins
somebrboosandorib when up.. Pads, brab in the vast incor..Pcwad
podirmef, let toesing..X;Himo,y Jencamizis anyYory, agotial.I'
ve the Anc
```

```
print(evaluate('Th', 200, temperature=0.8))
```

```
Thank it's can to get Pastion.
Think it make a prees incredisanted been Perice wreat? Eally s
ay soled a creaint things anything, if it.I don't im and actua
So here a lot jobs from going to get nest
```

```
print(evaluate('Th', 200, temperature=0.2))
```

```
Think your because the world we have and the world we have and
the world we have the world we have the was and the world we're
going to be and we have a lot of the world be the world we have
the said "W
```

At temperature=0.2, we can generate a sentence with real English words.

Furthermore, in my output in part 1, at `0m 10s` and `3m 16s`, it generates some blank lines. This might because in the original training set, there are many blank lines and blank spaces in the text file.

Note: The output you see from the jupyter notebook may not be the same as here, because I run the code for several times.

3. **Report perplexity on a couple validation texts that are similar and different to the training data.**

In order to avoid overflow/underflow error, we take log and then take exponential to perplexity:

Perplexity
= P(c1 c2 c3 ... cN) ^ (-1/N)
= exp (log (P(c1 c2 c3 ... cN) ^ (-1/N)))
= exp ((-1/N) *log (P(c1 c2 c3 ... cN))
= exp ((-1/N) * (log P(c1) + log P(c2|c1) + log P(c3|c1c2) + … + P(cN|c1c2…cN-1)))

In evaluation function, there is a variable output_dist that can help us to calculate P(c|prime_str). So I build a function **next_char_prob(c, prime_str)** that can compute P(c|prime_str) and then I build a function **perplexity(sentence)** that can calculate the perplexity of a sentence.

By setting temperature as 1.4, I generate a sentence

```
Wh', FHi8. It'se going Penksed P- the PHering care so cheae, in bild m
illions, if he's, I not bad. Vze
```

By setting temperature as 0.2, I generate a sentence

```
Where we have the world what we have the said the world we have the wor
ld be the startion and the worl
```

In my **next_char_prob** function, I set the temperature to 0.2, and then by using my **perplexity** function I get the perplexity for the first sentence above is 593.15 and the perplexity for the second sentence is 1.39 which is much smaller than the first one. This is true because the second sentence has more meaningful and real words than the first sentence.